Abstract

The content of a media program is recognized by analyzing its audio content to extract therefrom prescribed features, which are compared to a database of features associated with identified content. The identity of the content within the database that has features that most closely match the features of the media program being played is supplied as the identity of the program being played. The features are extracted from a frequency domain version of the media program by a) filtering the coefficients to reduce their number, e.g., using triangular filters; b) grouping a number of consecutive outputs of triangular filters into segments; and c) selecting those segments that meet prescribed criteria, such as those segments that have the largest minimum segment energy with prescribed constraints that prevent the segments from being too close to each other. The triangular filters may be log-spaced and their output may be normalized.